Peano Arithmetic

Goals Now

1) We will introduce a standard set of axioms for the language $\mathcal{L}_A$. The theory generated by these axioms is denoted **PA** and called Peano Arithmetic. Since **PA** is a sound, axiomatizable theory, it follows by the corollaries to Tarski's Theorem that it is incomplete. Nevertheless, it appears to be strong enough to prove all of the standard results in the field of number theory (including such things as the prime number theorem, whose standard proofs use analysis). Even Andrew Wiles' proof of Fermat's Last Theorem has been claimed to be formalizable in **PA**.

2) We know that **PA** is sound and incomplete, so there are true sentences in the language $\mathcal{L}_A$ which are not theorems of **PA**. We will outline a proof of Gödel's Second Incompleteness Theorem, which states that a specific true sentence, asserting that **PA** is consistent, is not a theorem of **PA**. This theorem can be generalized to show that any consistent theory satisfying general conditions cannot prove its own consistency.

3) We will introduce a finitely axiomatized subtheory **RA** ("Robinson Arithmetic") of **PA** and prove that every consistent extension of **RA** (including **PA**) is "undecidable" (meaning not recursive). As corollaries, we get a stronger form of Gödel's first incompleteness theorem, as well as Church's Theorem: The set of valid sentences of $\mathcal{L}_A$ is not recursive.

The Theory PA (Peano Arithmetic)

The so-called Peano postulates for the natural numbers were introduced by Giuseppe Peano in 1889. In modern form they can be stated in the language of set theory as follows. Let $\mathbb{N}$ be a set containing an element 0, and let $S : \mathbb{N} \to \mathbb{N}$ be a function satisfying the following postulates:

GP1: $S(x) \neq 0$, for all $x \in \mathbb{N}$.
GP2: If $S(x) = S(y)$ then $x = y$, for all $x, y \in \mathbb{N}$.
GP3: Let A be any subset of $\mathbb{N}$ which contains 0 and which is closed under S (i.e. $S(x) \in A$ for all $x \in A$). Then $A = \mathbb{N}$.

Note that GP3 is a form of induction.

It is not hard to show that any two systems $\langle \mathbb{N}, S, 0\rangle$ and $\langle \mathbb{N}', S', 0'\rangle$ which both satisfy GP1,GP2, GP3 are isomorphic, in the sense that there is a bijection $\phi : \mathbb{N} \to \mathbb{N}'$ such that $\phi(0) = 0'$ and

$$\phi(S(x)) = S'(\phi(x)), \text{ for all } x \in \mathbb{N}$$

Thus the Peano postulates characterize $\mathbb{N}$ up to isomorphism.

However, when it comes to designing a formal theory in the predicate calculus based on these Peano postulates we cannot formulate GP3 except in the context of formal set theory. It turns out to be essentially impossible to formulate a completely satisfactory theory of sets.

One simple solution is to design a "first-order" theory of $\mathbb{N}$ in which the universe is supposed to be $\mathbb{N}$ and the underlying language is $[0, s; =]$. This was done on pages 49-50, and the result is a complete theory Th(s) which can be completely axiomatized. However this theory cannot formulate much of interest, because $+$ and $\cdot$ cannot be defined in this language.

Thus to formulate our theory **PA** we extend this simple language by adding $+$ and $\cdot$ to obtain the language $\mathcal{L}_A = [0, s, +, \cdot; =]$. In this language, postulates GP1 and GP2 are easily formulated. The best we can do to formulate GP3 is to represent sets by formulas $A(x)$ in the language $\mathcal{L}_A$, where $A(x)$ is supposed to represent the set $\{x \mid A(x)\}$. When this is done carefully, we come up with the Induction Scheme below.

In order to complete the axioms of **PA** we need recursive definitions of $+$ and $\cdot$. These are formulated below as P3, P4 for $+$ and P5,P6 for $\cdot$.

Axioms for PA

P1 $\forall x(sx \neq 0)$
P2 $\forall x \forall y(sx = sy \supset x = y)$ s is 1-1 function
P3 $\forall x(x + 0 = x)$
P4 $\forall x \forall y(x + sy = s(x + y))$ } define $+$
P5 $\forall x(x \cdot 0 = 0)$
P6 $\forall x \forall y(x \cdot sy = (x \cdot y) + x)$ } define $\cdot$

Induction Scheme: Let $Ind(A(x))$ be the sentence

$$\forall y_1 \cdots \forall y_k[(A(0) \wedge \forall x(A(x) \supset A(sx))) \supset \forall x A(x)]$$

where A is any formula whose free variables are among $x, y_1, \cdots, y_k$. (The variables $y_1, \cdots, y_k$ are called parameters.) All such sentences $Ind(A)$ are axioms.

Let $\Gamma_{PA} = \{P_1, \ldots, P_6\} \cup \{\text{Induction axioms}\}$. Then Γ_{PA} is recursive. This is clear from Church's thesis.

Definition: $\mathbf{PA} = \{A \in \Phi_0 \mid \Gamma_{PA} \models A\}$

Thus **PA** is an axiomatizable theory. It is a sound theory since all of its axioms (and hence all of its theorems) are true in the standard model $\underline{\mathbb{N}}$.

Terminology: We speak of sentences in **PA** as *theorems* of **PA**, because they can be proved (for example, by LK proofs), from the axioms of **PA**. We use the notation $\mathbf{PA} \vdash A$ to mean that A is a theorem of **PA**.

Example 1:
We show that **PA** proves that all nonzero elements have predecessors. Let

$$A(x) \;=\; (x = 0 \vee \exists y(x = sy))$$

In order to prove this by induction there are two steps:

Basis: $x = 0$ $\qquad$ $\mathbf{PA} \vdash A(0)$

Induction Step: $z \leftarrow sz$ $\qquad$ $\mathbf{PA} \vdash \forall x(A(x) \supset A(sx))$

In fact, both $A(0)$ and $\forall x(A(x) \supset A(sx))$ are valid sentences, so no axioms of **PA** are needed to show that they are theorems of **PA**. It follows from the induction axiom $Ind(A(x))$ that

$$\mathbf{PA} \vdash \forall x A(x)$$

Example 2:
We show that **PA** proves the associative law for +. Let

$$A(z) \;=\; (x + y) + z = x + (y + z)$$

We use the induction axiom $Ind(A(z))$.

Basis: $z = 0$

$$\begin{aligned} (x + y) + 0 &= x + y && \text{P3} \\ &= x + (y + 0) && \text{P3} \end{aligned}$$

Induction Step: $z \leftarrow sz$

$$\begin{aligned} (x + y) + sz &= s((x + y) + z) && \text{P4} \\ &= s(x + (y + z)) && \text{Induction Hypothesis} \\ &= x + s(y + z) && \text{P4} \\ &= x + (y + sz) && \text{P4} \end{aligned}$$

Thus by $Ind(A(z))$ it follows that

$$\mathbf{PA} \vdash \forall x \forall y \forall z A(z)$$

Exercise 1 *Show that* **PA** *proves the commutative law of addition, the associative and commutative laws of multiplication, and that multiplication distributes over addition, using the style of Example 2. In each case state carefully which induction axiom (or axioms) are needed, and which axioms P1,...,P6 are needed, (or which earlier results).*

Exercise 2 *Recall the theory of successor $Th(s)$ presented on pages 49-50. Show that all of the axioms S3, S4, S5, ... follow from S1 and S2 together with the Induction Scheme $Ind(A(x))$ for all formulas $A(x)$ in the language of successor $[0, s; =]$.*

PA is incomplete, because it is axiomatizable and sound (and has $\mathcal{L}_A$ as the underlying language): see Corollary 3, page 95. Later we will give explicit true sentences that are not theorems of **PA**, including the assertion that **PA** is consistent.

An apparent paradox is that the Peano postulates GP1, GP2, GP3 characterize the natural numbers in set theory (as explained above), and yet there are nonstandard models for **PA**. (We know there are nonstandard models both from the fact that **PA** is incomplete, and by the construction using compactness given on page 51.) However, the Peano Axioms only characterize the natural numbers under the assumption that we could do induction using an arbitrary set. In **PA**, we can only use induction on arithmetical sets.

Observed fact: All standard theorems of number theory are in **PA**. Even Wiles' 1995 proof of "Fermat's Last Theorem" apparently can be formalized in **PA**. So famous open problems, such as Goldbach's conjecture and the prime pair conjecture, can probably be either proved or disproved in **PA**. Goldbach's conjecture can certainly be disproved in **PA** if it is false: just present and verify a counter example. (Is the same true for the prime pair conjecture?)

RA: A finitely axiomatized subtheory of PA

Our main tool for showing that a theory such as **PA** is undecidable is showing that every r.e. relation (including the undecidable set K) is representable in the theory (see the definition below). This argument applies not only to **PA**, but to a weak subtheory of **PA** known as **RA**.

Recall the syntactic definition of $\leq$ given on page 84: $t_1 \leq t_2$ stands for $\exists z(t_1 + z = t_2)$, where z is a new variable.

We now extend P1,...,P6 with three new axioms.

P7 $\forall x(x \leq 0 \supset x = 0)$
P8 $\forall x \forall y(x \leq sy \supset (x \leq y \vee x = sy))$
P9 $\forall x \forall y(x \leq y \vee y \leq x)$

Definition: **RA** is the theory whose axioms are P1, $\cdots$, P6, P7, P8, P9.

Note that **RA** has no induction axioms. We note three important facts about **RA**:

1) **RA** $\subseteq$ **PA** (i.e. P7, P8, P9 are in **PA** because they can be proved by induction).

2) **RA** has only finitely many axioms.

2) The axioms of **RA** are $\forall$-sentences (over $\mathcal{L}_{A,\leq}$).

Later we will show that **RA** $\neq$ **PA**.

Exercise 3 *Show that P7, P8, P9 are each theorems of* **PA**. *First translate each axiom into the language $\mathcal{L}_A$ by getting rid of $\leq$ (see page 84).*

Definition: A theory Σ is *decidable* iff $\{\#A | A \in \Sigma\}$ is recursive.

Informally, Σ is decidable iff there is an algorithm which, given any sentence A, determines whether A is in Σ.

Definition: If Σ and Σ' are theories, then Σ' is an *extension* of Σ if $\Sigma \subseteq \Sigma'$.

We will show that **RA** is undecidable, and use this to prove that in fact every sound theory (over the language $\mathcal{L}_A$) is undecidable. Our main tool is the representation theorem below. Recall the definition (bottom of page 83) for a formula $A(\vec{x})$ to represent a relation $R(\vec{x})$. We now extend this definition to apply to a theory Σ.

Definition: A formula $A(\vec{x})$ *represents* a relation $R(\vec{x})$ in a theory Σ if for all $\vec{a} \in \mathbb{N}^n$

$$R(\vec{a}) \Leftrightarrow \Sigma \vdash A(s_{\vec{a}})$$

Note that according to our earlier definition, $A(\vec{x})$ represents $R(\vec{x})$ (with no theory mentioned) iff $A(\vec{x})$ represents $R(\vec{x})$ in **TA**.

Recall the definition (page 85) of a $\exists\Delta_0$ formula.

RA Representation Theorem: Every r.e. relation is representable in **RA** (and in every sound extension of **RA**) by an $\exists\Delta_0$ formula.

This is a major result and will take several pages to prove. Of course we already know from the Exists Delta Theorem (page 86) that every r.e. relation is representable in **TA**. The extra work now is showing that the true $\exists\Delta_0$ formulas are provable in **RA**.

Before giving the proof of the Theorem, we prove several consequences.

Corollary 1: Every sound extension of **RA** (including **PA**) is undecidable.

Proof: Let Σ be a sound extension of **RA**. It suffices to show $K \leq_m \Sigma$, or more precisely to show that $K \leq_m \hat{\Sigma}$, where $\hat{\Sigma}$ is the set of codes for theorems of Σ; that is $\hat{\Sigma} = \{\#A \mid \Sigma \vdash A\}$ (see page 90).

Since K is r.e., it follows from the theorem that K is represented in Σ by some $\exists\Delta_0$ formula $A(x)$. Thus for all $a \in \mathbb{N}$

$$a \in K \Leftrightarrow \Sigma \vdash A(s_a)$$

Define the total computable function $f : \mathbb{N} \to \mathbb{N}$ by

$$f(a) = \#A(s_a)$$

Then f is clearly computable by Church's thesis, and in fact it can be obtained from the computable function $\text{sub}(m, n)$ defined on page 91. Namely, $f(a) = \text{sub}(m_0, a)$, where $m_0 = \#A(x)$. Thus

$$a \in K \Leftrightarrow f(a) \in \hat{\Sigma}$$

as required. □

Recall Corollary 1, page 94 states that the set $VALID$ of valid sentences of $\mathcal{L}_A$ is r.e. Now we can prove more:

Corollary 2: Church's Theorem: The set $VALID$ of valid sentences in the language $\mathcal{L}_A$ is undecidable.

Proof: We use the fact that **RA** is undecidable, and has only finitely many axioms, P1,...,P9. Let γ be the conjunction P1$\wedge$...$\wedge$P9 of these axioms. Then

$$A \in \mathbf{RA} \iff (\gamma \supset A) \text{ is valid}$$

Hence we've reduced the problem of membership in **RA** to the validity problem, so validity is undecidable. (We've only given an informal argument for the reduction, so we need Church's thesis here.) $\square$

Remark: In fact, the validity problem is undecidable for any language that contains a binary predicate symbol. This can proved directly by reduction of the halting problem for Turing machines to validity, as was shown in Turing's famous 1936 paper introducing Turing machines.

Decidability Theorem: Every complete axiomatizable theory is decidable.

Proof: We give an informal proof, using Church's thesis. If Σ is axiomatizable, then by the theorem on page 93, it is r.e. Here is an algorithm for determining whether a given formula A is in Σ, assuming that Σ is complete. Enumerate the members of Σ. Sooner or later, either A or $\neg A$ will appear in the enumeration. If A appears, then it is in Σ. If $\neg A$ appears, then A is not in Σ. $\square$.

Now we can obtain an alternative proof of Corollary 3 to Tarski's Theorem, page 95:

Corollary: Every sound axiomatizable theory is incomplete.

Proof: Let Σ be a sound axiomatizable theory. If Σ is not an extension of **RA** it is certainly incomplete. If Σ is an extension of **RA**, then by Corollary 1 above Σ is undecidable, and hence by the Decidability Theorem Σ is incomplete. $\square$

Exercise 4 *Prove that there is an $\exists\Delta_0$ sentence A such that $\neg A \in$ **TA** but **PA** $\nvdash \neg A$. (Compare this with Corollary 2, page 104.)*

In order to prove the **RA** Representation Theorem we need to recall the syntactic definitions involving $\leq$ given on page 84.

MAIN LEMMA: Every bounded sentence in **TA** is in **RA**. That is, every true bounded sentence can be proved from the axioms of **RA**. (Thus $\mathbf{TA} \cap \Delta_0 = \mathbf{RA} \cap \Delta_0$.)

Notation: When we write a specific number such as 4 in an example formula, this is an abbreviation for the corresponding numeral; s_4 (i.e. $ssss0$) in this case.

Example of a true bounded sentence:

$$\forall x \leq 1000 \exists y \leq 2 \cdot x[x = 0 \vee (x < y \wedge \text{Prime}(y))]$$

Notice that since the quantifiers are bounded, and the assertion is being made for only finitely many pairs x, y. Each case can be proved separately by "brute force".

To prove the MAIN LEMMA it is easier to expand the language $\mathcal{L}_A$ to $\mathcal{L}_{A,\leq}$ by adding the binary connective $\leq$ as a primitive symbol (see page 84). Then we expand the theory **RA** to the theory $\mathbf{RA}_\leq$ over the language $\mathcal{L}_{A,\leq}$ by interpreting $\leq$ in the axioms P7,P8,P9 as a primitive symbol, and by adding the new axiom

P0 $\forall x \forall y (x \leq y \leftrightarrow \exists z (x + z = y))$

Every formula A over $\mathcal{L}_{A,\leq}$ can be translated to a formula A' over $\mathcal{L}_A$ by replacing each atomic subformula of the form $t_1 \leq t_2$ in A by the formula $\exists z(t_1 + z = t_2)$, where z is a variable not occurring in t_1, t_2 (see page 84). Notice that if $\leq$ does not occur in A, then $A =_{syn} A'$.

Translation Lemma: For every formula A over $\mathcal{L}_{A,\leq}$,

$$\mathbf{RA}_\leq \vdash A \text{ iff } \mathbf{RA} \vdash A'$$

Proof: There is a natural one-one correspondence between models of $\mathbf{RA}_\leq$ and **RA**, namely for each model $\mathcal{M}$ of **RA** we associate the model $\hat{\mathcal{M}}$ of $\mathbf{RA}_\leq$ which is the same as $\mathcal{M}$ except we add the interpretation of $\leq$ in such a way that axiom P0 is satisfied. Then we claim that for every $\mathcal{L}_{A,\leq}$ formula A

$$\hat{\mathcal{M}} \models A \text{ iff } \mathcal{M} \models A'$$

The claim is easily proved by structural induction on A. The Translation Lemma follows easily from the claim. □

Proof of MAIN LEMMA: We prove the MAIN LEMMA for $\mathbf{RA}_\leq$. It follows for **RA** by the Translation Lemma.

Let A be a true bounded sentence. Move all $\neg$'s in A past other connectives so that they govern only atomic formulas $t = u$. Do this by using DeMorgan's Laws, and the equivalences

$$\neg\neg A \Longleftrightarrow A, \quad \neg\forall x \leq t\ B \Longleftrightarrow \exists x \leq t \neg B, \quad \neg\exists x \leq t\ B \Longleftrightarrow \forall x \leq t \neg B$$

Exercise 5 *Show from the definitions of the bounded quantifiers $\exists x \leq t$ and $\forall x \leq t$ that for each of the three equivalences above the formulas on the left and right are logically equivalent (this is obvious for the first equivalence).*

The proof of the MAIN LEMMA proceeds by induction on the number of logical operators (other than $\neg$) in this modified A.

For the base case, A has one of the four forms $t = u$, $t \neq u$, $t \leq u$, $\neg t \leq u$.

Example: A is $s0 + s0 = ss0$. This can be proved in **RA** by the recursive definition of $+$:

$x + 0 = x$ (P3)
$x + sy = s(x + y)$ (P4)

More generally:

Lemma A1: For all $m, n \in \mathbb{N}$,

$$\mathbf{RA} \vdash s_m + s_n = s_{m+n} \text{ and}$$
$$\mathbf{RA} \vdash s_m \cdot s_n = s_{m \cdot n}$$

Proof: The first line is proved by induction (outside the system) on n using P3 and P4, as in the example. Then the second line is proved by induction on n using P5, P6, and the first line. □

If t is any closed term (i.e. with no variables), then $t^{\mathcal{M}} = n$ for some $n \in \mathbb{N}$, where $\mathcal{M}$ is the standard model. Thus $t = s_n \in \mathbf{TA}$.

Lemma A: If t is a closed term and $t = s_n$ is in **TA**, then $\mathbf{RA} \vdash t = s_n$.

Proof: Induction on the length of t, using Lemma A1.

Lemma B: If $m < n$, then $\mathbf{RA} \vdash s_n \neq s_m$.

Proof: Induction on m, using P1 and P2. □

For example, consider $ss0 \neq s0$. Recall that P2 is $\forall x(sx = sy \supset x = y)$. Thus $ss0 = s0 \supset s0 = 0$. But by P1, $s0 \neq 0$. Therefore $ss0 \neq s0$.

Remark: Arguments such as the one above could be formalized by an LK proof using the equality axioms. However the implications are clear without bothering to carry out such a formal proof, if we keep in mind the definition of logical consequence (page 23), and the Basic Semantic Definition (page 22), and in particular that $=$ must be interpreted as equality in any structure.

The base case for the MAIN LEMMA for the sentences $t = u$ and $t \neq u$ follows easily from Lemma A and Lemma B. For the case $t \leq u$ we apply P0, so the problem reduces to the first case of Lemma A1. The case $\neg t \leq u$ follows from Lemma C below, together with Lemma B.

The induction step for the MAIN LEMMA follows from the following:

Lemma C: For all n, $\mathbf{RA}_{\leq}$ proves the sentence

$$\forall x(x \leq s_n \supset (x = 0 \vee x = s_1 \vee \ldots \vee x = s_n))$$

Proof: Induction on n. The base case is $x \leq 0 \supset x = 0$, which is P7. The induction step follows easily from P8. □ (Lemma C)

For the induction step in the proof of the MAIN LEMMA, let A be a true bounded sentence. We assume that $\neg$'s in A have been driven in as explained above, and A does not fit the base case, so the principle connective of A is one of $\wedge$, $\vee$, $\forall \leq$, $\exists \leq$. The cases of $\wedge$ and $\vee$ are trivial: just apply the induction hypothesis.

Now consider the case $\forall \leq$, say A is $\forall x \leq tB(x)$, and this is in **TA**. Since this is a sentence, and by definition of $\forall x \leq t$, x cannot occur in t, it follows that t is a closed term. Thus by Lemma A, **RA** can prove $t = s_n$ for some n.

For example, suppose $n = 23$. Then it suffices to show that $\forall x \leq 23\ B(x)$ is provable in $\mathbf{RA}_{\leq}$. By Lemma C, $\mathbf{RA}_{\leq}$ proves

$$x \leq 23 \supset (x = 0 \vee x = 1 \vee \cdots \vee x = 23)$$

By the Substitution Theorem (page 26) it follows in general, that for any closed term u,

$$\forall x(x = u \supset (B(u) \leftrightarrow B(x)))$$

is valid. Therefore it follows by reasoning in $\mathbf{RA}_{\leq}$ that $\forall x \leq tB(x)$ is implied by

$$B(0) \wedge B(1) \wedge \cdots \wedge B(23)$$

Since $\forall x \leq tB(x)$ is true, it follows that $B(0), B(1), \ldots$ are each true, so by the induction hypothesis each is in $\mathbf{RA}_{\leq}$. Hence their conjunction is in $\mathbf{RA}_{\leq}$, so $\forall x \leq tB$ is in $\mathbf{RA}_{\leq}$.

The case $\exists \leq$ is easier than the $\forall \leq$ case and does not require Lemma C. □ (MAIN LEMMA)

Exercise 6 *Prove the $\exists \leq$ case in the above proof.*

Corollaries to MAIN LEMMA

Corollary 1: The set of bounded sentences of **TA** is decidable. (This can also be proved without the MAIN LEMMA, as was intended in Exercise 6, page 91.)

Corollary 2: Every $\exists\Delta_0$ sentence (page 85) of **TA** is provable in **RA**.

Corollary 3: The set of $\exists\Delta_0$ sentences of **TA** is r.e. (but not decidable).

Exercise 7 *Prove the above three corollaries.*

Exercise 8 *Let $\exists y A(x, y)$ be a $\exists\Delta_0$ formula which represents $K(x)$ in* **RA** *(where $K(x) = (\{x\}_1(x) \neq \infty)$ is the standard halting problem). Show that there is a consistent extension Σ of* **RA** *such that $\exists y A(x, y)$ does not represent $K(x)$ in Σ.* **Hint:** *Form Σ by adding a suitable false axiom to* **RA** *which retains consistency.*

Proof of RA Representation Theorem: (See page 100 for the statement.)

Proof: Suppose $R(\vec{x})$ is an r.e. relation. By the Exists Delta Theorem (page 86) $R(\vec{x})$ is represented in **TA** by some $\exists\Delta_0$ formula $\exists y A(\vec{x}, y)$. Thus for all $\vec{a} \in \mathbb{N}^n$,

$$R(\vec{a}) \Leftrightarrow [\exists y A(s_{a_1}, ..., s_{a_n}, y) \in TA]$$

By Corollary 2 above and the soundness of **RA**, this is equivalent to

$$R(\vec{a}) \Leftrightarrow [\Sigma \vdash \exists y A(s_{a_1}, ..., s_{a_n}, y)]$$

where Σ is any sound extension of **RA**(i.e. **RA** $\subseteq \Sigma \subseteq$ **TA**). Thus by the definition $\exists y A(\vec{x}, y)$ represents $R(\vec{x})$ in Σ. □

The following is a generalization of Church's Theorem (page 101).

Theorem: Every sound theory is undecidable.

Exercise 9 *Prove the theorem.*

Results for consistent (possibly unsound) theories

Our goal now is to prove the following theorem:

Main Theorem: Every consistent extension of **RA** is undecidable.

Corollary: Every consistent axiomatizable extension of **RA** is incomplete.

Proof of Corollary: This follows from the Decidability Theorem (page 101). □

Notice that this strengthens the Corollary 3, page 95, to Tarski's Theorem, since we no longer need to assume soundness in order to conclude that an axiomatizable theory is incomplete (provided that the theory includes **RA**). Notice that soundness is a semantic notion, whereas consistency can be given a syntactic definition (there is no proof of 0=1). The proof of the Main Theorem can be made to avoid the complex semantic notion of truth of an arbitrary sentence of $\mathcal{L}_A$.

An example of an unsound consistent extension of **RA** is the theory $Th(\mathbb{Z}[X]^+)$ consisting of all sentences in the language $\mathcal{L}_A$ which are true in the structure $\mathbb{Z}[X]^+$, where the universe of $\mathbb{Z}[X]^+$ is the set of all polynomials $p(X)$ with integer coefficients such that either $p(X)$ is the zero polynomial, or the leading coefficient of $p(X)$ is positive. (Here $+$ and $\cdot$ are polynomial addition and multiplication, and the successor of $p(X)$ is $p(X) + 1$.) The axioms P1,...,P9 are in the theory $Th(\mathbb{Z}[X]^+)$, but the theory is unsound, because the sentence

$$A = \quad \exists x \forall y (x \neq y + y \wedge x \neq y + y + s0) \tag{1}$$

is not in **TA** but is in $Th(\mathbb{Z}[X]^+)$. (To check the latter claim, let x be the polynomial X.)

Thus $Th(\mathbb{Z}[X]^+)$ is undecidable, by the Main Theorem.

Corollary: RA$\neq$ PA

Proof: Let A be the sentence in (1) above. Then $\neg A$ is a theorem of **PA** (it can be proved by induction on x), but $\neg A$ is not a theorem of **RA**, since the structure $\mathbb{Z}[X]^+$ just described is a model of **RA** which satisfies A.

Exercise 10 *Is $Th(\mathbb{Z}[X]^+)$ axiomatizable? Justify your answer.*

Notice that the structure $\mathbb{Z}[X]^+$ is a nonstandard model of **RA**. There are no such nice nonstandard models of **PA**. In fact one can prove that for any nonstandard model of **PA** with universe $\mathbb{N}$, the interpretations of $+$ and $\cdot$ are uncomputable functions.

In order to prove the Main Theorem we need a stronger notion of representability.

Recall the definition of *represents in a theory* Σ (page 100):

A represents R in Σ iff $\forall \vec{a} \in \mathbb{N}^n$ $(R(\vec{a}) \Leftrightarrow A(s_{\vec{a}}) \in \Sigma)$.

Definition: *A strongly represents R* in Σ iff $\forall \vec{a} \in \mathbb{N}^n$
$R(\vec{a}) \Rightarrow (A(s_{\vec{a}}) \in \Sigma)$, and $\neg R(\vec{a}) \Rightarrow (\neg A(s_{\vec{a}}) \in \Sigma)$

Notice that if Σ is a consistent theory, then if $A(\vec{x})$ strongly represents $R(\vec{x})$ in Σ it follows that $A(\vec{x})$ also represents $R(\vec{x})$ in Σ. The converse is not always true (unless Σ is complete).

In order to prove the Main Theorem, we will prove the following two results:

Undecidability Theorem: If every recursive relation is representable in a theory Σ then Σ is undecidable.

Strong RA Representation Theorem: Every recursive relation is strongly representable in **RA** by an $\exists\Delta_0$ formula.

Exercise 11 *Prove the converse of the above Theorem: If R is strongly representable in* **RA***, then R is recursive.*

Proof of the Main Theorem: This follows from the preceding two theorems by the following simple fact: If a relation is strongly representable in **RA** then it is strongly representable in every extension of **RA**, and hence it is representable (rather than strongly representable) in every consistent extension of **RA**. This is immediate from the definitions of representable and strongly representable (page 106). □

We now turn to the proof of the Undecidability Theorem. First note that if the hypothesis of this theorem is strengthened to assume that every r.e. (as opposed to recursive) relation is representable in Σ, then it would be very easy to prove that Σ is undecidable. (See the proof of Corollary 1 to the **RA** Representation Theorem, page 100). The reason the theorem is stated with the weaker hypothesis is to make the argument in the preceding paragraph work. See exercise 8 to see what goes wrong when using the alternative form of the Undecidability Theorem.

Proof of the Undecidability Theorem: (Like the proof of Tarski's Theorem)

Assume Σ is recursive. The idea is to formulate a sentence "I am not in Σ". This should be true, because Σ is consistent, but then it should be in Σ by representability, a contradiction.

Recall $d(x) = \text{sub}\ (x, x)$ from the proof of Tarski's theorem. Then d is a function (semantic notion) with the property that for all $a \in \mathbb{N}$, $d(a) = \#A(s_a)$ where $a = \#A(x)$. Note that d is a recursive function.

Define $R(x) \Leftrightarrow (x = \#A$, for some $A \in \Sigma)$. Thus $R = \hat{\Sigma}$, and Σ is recursive iff R is recursive. In order to get a contradiction, assume R is recursive. Let

$$S(x) \Leftrightarrow \neg R(d(x))$$

Then S is recursive. Hence by hypothesis, $S(x)$ is represented in Σ by some formula $B(x)$.

By definition of representable
(1) $\quad \neg R(d(a)) \Leftrightarrow (B(s_a) \in \Sigma), \quad$ for all $a \in \mathbb{N}$

Let $e = \#B(x)$. Then $d(e) = \#B(s_e)$ by definition of $d(x)$. Then by (1),

$$\neg R(d(e)) \Leftrightarrow (B(s_e) \in \Sigma)$$

The LHS asserts $B(s_e) \notin \Sigma$, because R represents membership in Σ. This is a contradiction, hence Σ is not recursive. $\square$

Proof of the Strong RA Representation Theorem: Suppose $R(\vec{x})$ is a recursive relation. Then both R and $\neg R$ are r.e., so by the Exists Delta Theorem, there are bounded formulas B_1 and B_2 such that $\exists y B_1(\vec{x}, y)$ represents $R(\vec{x})$ in **TA** and $\exists y B_2(\vec{x}, y)$ represents $\neg R(\vec{x})$ in **TA**. As pointed out in the previous proof, $\exists y B_1(\vec{x}, y)$ also represents $R(\vec{x})$ in **RA**, but in general it will not strongly represent $R(\vec{x})$ in **RA**. For strong representation we define a formula

$$A(\vec{x}) \equiv \exists y[B_1(\vec{x}, y) \wedge \forall z \leq y \neg B_2(\vec{x}, z)]$$

Claim: $A(\vec{x})$ strongly represents $R(\vec{x})$ in **RA**.

First we establish that for all $\vec{a} \in \mathbb{N}$,

$$R(\vec{a}) \Rightarrow \mathbf{RA} \vdash A(s_{\vec{a}}) \tag{2}$$

Since $\exists y B_1(\vec{x}, y)$ represents $R(\vec{x})$ in **RA**, we conclude from $R(\vec{a})$ that

$$\mathbf{RA} \vdash B_1(s_{\vec{a}}, s_b), \text{ for some } b \in \mathbb{N}$$

By the property of B_2 we know $\forall z \leq s_b \neg B_2(s_{\vec{a}}, z) \in \mathbf{TA}$, so by the MAIN LEMMA this sentence is in **RA**. This establishes (2) (take $y = b$).

It remains to establish

$$\neg R(\vec{a}) \Rightarrow \mathbf{RA} \vdash \neg A(s_{\vec{a}}) \tag{3}$$

Assume $\neg R(\vec{a})$. Note that $\neg A(s_{\vec{a}})$ is equivalent to

$$\forall y[\neg B_1(s_{\vec{a}}, y) \vee \exists z \leq y B_2(s_{\vec{a}}, z)] \tag{4}$$

Since $\exists z B_2(\vec{x}, z)$ represents $\neg R(\vec{x})$ in **RA** it follows that for some $c \in \mathbb{N}$

$$\mathbf{RA} \vdash B_2(s_{\vec{a}}, s_c) \tag{5}$$

By P9,

$$\mathbf{RA} \vdash \forall y(y \leq s_c \vee s_c \leq y)$$

(This is the only place that P9 is needed.) Thus to establish (4) in **RA** we consider two cases, depending on whether $y \leq s_c$ or $s_c \leq y$. For the first case, we note that

$$\forall y \leq s_c \neg B_1(s_a, y)$$

is a true bounded formula, and therefore by the MAIN LEMMA provable in **RA**, so (4) follows in **RA**

For the second case, by (5) we have

$$\mathbf{RA} \vdash \forall y(s_c \leq y \supset \exists z \leq y B_2(s_a, z))$$

so again (4) follows in **RA**. □

Exercise 12 *Let* $\neg\mathbf{RA} = \{A \mid \mathbf{RA} \vdash \neg A\}$*. Thus* $\neg\mathbf{RA}$ *is the set of sentences which* **RA** *proves false. Prove that* **RA** *and* $\neg\mathbf{RA}$ *are recursively inseparable. That is, prove that there is no recursive set* S *of sentences such that*

$$\mathbf{RA} \subseteq S \text{ and } \neg\mathbf{RA} \subseteq S^c$$

where $S^c = \{A \in \Phi_0 \mid A \notin S\}$*. (Note that* S *need not be a theory.)*

Hint: *Study the proof of Tarski's Theorem (page 91) and of the Undecidability Theorem (page 107). Assume that there is a recursive set* S *satisfying the indicated conditions. Formulate a sentence asserting "I am not in* S*", and obtain a contradiction.*